Congratulations! You passed!

Grade received 80% **Latest Submission Grade** 80% **To pass** 80% or higher

Go to next item

1. What is the "cache" used for in our implementation of forward propagation and backward propagation? **1 / 1 point**

- ○ It is used to cache the intermediate values of the cost function during training.
- ○ We use it to pass variables computed during backward propagation to the corresponding forward propagation step. It contains useful values for forward propagation to compute activations.
- ○ It is used to keep track of the hyperparameters that we are searching over, to speed up computation.
- ◉ We use it to pass Z computed during forward propagation to the corresponding backward propagation step. It contains useful values for backward propagation to compute derivatives.

Expand

✓ Correct
Correct, the "cache" records values from the forward propagation units and are used in backward propagation units because it is needed to compute the chain rule derivatives.

2. During the backpropagation process, we use gradient descent to change the hyperparameters. True/False? **1 / 1 point**

- ◉ False
- ○ True

Expand

✓ Correct
Correct. During backpropagation, we use gradient descent to compute new values of $W^{[l]}$ and $b^{[l]}$. These are the parameters of the network.

3. Which of the following statements is true? **1 / 1 point**

- ◉ The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.
- ○ The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.

Expand

✓ Correct

4. Vectorization allows you to compute forward propagation in an L-layer neural network without an explicit for-loop (or any other explicit iterative loop) over the layers l=1, 2, …,L. True/False? **1 / 1 point**

- ◉ False
- ○ True

Expand

✓ Correct
Forward propagation propagates the input through the layers, although for shallow networks we may just write all the lines ($a^{[2]} = g^{[2]}(z^{[2]}), z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}$, ...) in a deeper network, we cannot avoid a for loop iterating over the layers: ($a^{[l]} = g^{[l]}(z^{[l]}), z^{[l]} = W^{[l]}a^{[l-1]} + b^{[l]}$, ...).

5. Suppose W[i] is the array with the weights of the i-th layer, b[i] is the vector of biases of the i-th layer, and g is the activation function used in all layers. Which of the following calculates the forward propagation for the neural network with L layers.

0 / 1 point

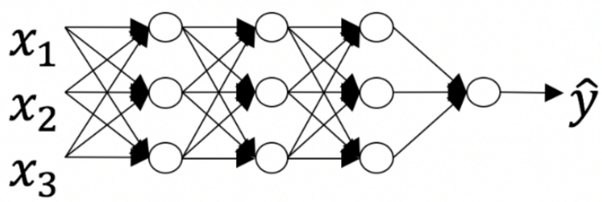
- (selected)
```
for i in range(1, L):
Z[i] = W[i]*A[i-1] + b[i]
A[i] = g(Z[i])
```
-
```
for i in range(L):
Z[i] = W[i]*X + b[i]
A[i] = g(Z[i])
```
-
```
for i in range(L):
Z[i+1] = W[i+1]*A[i+1] + b[i+1]
A[i+1] = g(Z[i+1])
```
-
```
for i in range(1, L+1):
Z[i] = W[i]*A[i-1] + b[i]
A[i] = g(Z[i])
```

Expand

Incorrect
No. Remember that the range omits the last number thus the range from 1 to L calculates only the A up to the L-1 layer.

6. Consider the following neural network.

1 / 1 point



How many layers does this network have?

- The number of layers L is 3. The number of hidden layers is 3.
- (selected) The number of layers L is 4. The number of hidden layers is 3.
- The number of layers L is 5. The number of hidden layers is 4.
- The number of layers L is 4. The number of hidden layers is 4.

Expand

Correct
Yes. As seen in lecture, the number of layers is counted as the number of hidden layers + 1. The input and output layers are not counted as hidden layers.

7. During forward propagation, in the forward function for a layer l you need to know what is the activation function in a layer (sigmoid, tanh, ReLU, etc.). During backpropagation, the corresponding backward function also needs to know what is the activation function for layer l, since the gradient depends on it. True/False?

1 / 1 point

- (selected) True
- False

Expand

Correct
Yes, as you've seen in week 3 each activation has a different derivative. Thus, during backpropagation you need to know which activation was used in the forward propagation to be able to compute the correct derivative.

8. There are certain functions with the following properties:

1 / 1 point

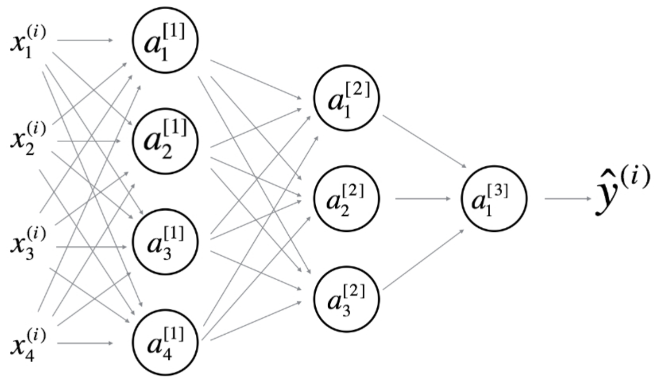
(i) To compute the function using a shallow network circuit, you will need a large network (where we measure size by the number of logic gates in the network), but (ii) To compute it using a deep network circuit, you need only an exponentially smaller network. True/False?

- (selected) True
- False

Expand

Correct

9. Consider the following 2 hidden layer neural network: 0 / 1 point



Which of the following statements are True? (Check all that apply).

- [x] $b^{[3]}$ will have shape (1, 1)

Correct

Yes. More generally, the shape of $b^{[l]}$ is $(n^{[l]}, 1)$.

- [ ] $W^{[3]}$ will have shape (3, 1)
- [x] $b^{[1]}$ will have shape (4, 1)

Correct

Yes. More generally, the shape of $b^{[l]}$ is $(n^{[l]}, 1)$.

- [ ] $W^{[3]}$ will have shape (1, 3)
- [ ] $W^{[2]}$ will have shape (3, 1)
- [ ] $b^{[3]}$

will have shape (3, 1)
- [ ] $$W^{[1]}$$ will have shape (4, 4)
- [ ] $$W^{[1]}$$ will have shape (3, 4)
- [x] $$W^{[2]}$$ will have shape (3, 4)

Correct

Yes. More generally, the shape of $$W^{[l]}$$ is $$(n^{[l]}, n^{[l-1]})$$.

- [ ] $$b^{[2]}$$ will have shape (1, 1)
- [x] $$b^{[2]}$$ will have shape (3, 1)

Correct

Yes. More generally, the shape of $$b^{[l]}$$ is $$(n^{[l]}, 1)$$.

Expand

Incorrect

You didn't select all the correct answers

10. In the general case if we are training with m examples what is the shape of $A^{[l]}$? 1 / 1 point

- ( ) $(m, n^{[l+1]})$
- ( ) $(n^{[l+1]}, m)$
- (x) $(n^{[l]}, m)$
- ( ) $(m, n^{[l]})$

Expand

Correct

Yes. The number of rows in $A^{[1]}$ corresponds to the number of units in the l-th layer.